图书情报工作
第65卷 第15期 2021年8月

AI + 智慧知识服务生态体系研究设计与应用实践*

——以中国科学院文献情报中心智慧服务平台建设为例

■ 钱力[1,2] 刘细文[1,2] 张智雄[1,2] 刘会洲[1,2]

[1] 中国科学院文献情报中心 北京 100190

[2] 中国科学院大学经济与管理学院图书情报与档案管理系 北京 100190

摘 要：[目的/意义] 人工智能(AI)正引发链式反应般的科学突破，引领新一轮科技革命和产业变革，图书文献情报领域如何利用AI技术提供智慧知识服务与智能情报系统是当前行业关注的焦点与热点。[方法/过程] 从图书情报行业内外综合分析AI技术与大数据为知识服务范式带来的新平台、新服务以及新机遇与新挑战，提出“AI技术+大数据”驱动的智慧知识服务生态体系建设的总体思路，从智慧数据、智慧中台与智慧服务3个层面共同构建“科情大脑”，提供覆盖科技管理、科技创新与社会学术信息环境的开放智慧知识服务生态环境。[结果/结论] 以中国科学院文献情报中心的文献情报数据湖、智能知识服务引擎、智慧知识发现、智慧知识管理、智能情报分析系统以及智能感知环境6个方面进行探索建设，取得有意义的成效。面向未来，阐明AI技术在面向大数据治理、细粒度知识识别、精准服务提供等方面，仍需要在数据、技术以及服务模式上进一步提升。

关键词：人工智能(AI) 科技大数据 智慧知识服务 深度学习

分类号：G25

DOI：10.13266/j.issn.0252-3116.2021.15.010

1 引言

目前，人工智能正引发链式反应般的科学突破，引领新一轮科技革命和产业变革，而作为支撑人工智能发展的科技大数据，记载着科学真理验证过程、实验观测、研究结论、网络交流等科技情报知识线索，是人工智能用于科技创新发现的算法模型实现的数据根基和知识基础，而且以语义化知识数据为基础的知识服务及其“人-机-物”三元计算体系，已经成为谷歌、微软等企业抢占未来大数据人工智能服务的重要部署。在此基础上，基于语义数据的科技知识的深入挖掘和重构，促进了研究前沿识别、颠覆性技术识别和技术交叉前沿发现等科技创新发现，同时也成为不断丰富化、细粒度化和语义化的知识体系的一部分[1]。柯平教授等提出后知识服务时代以智慧服务为核心，使“数据(D)-智慧(W)”的智慧链条成为可能，强调了以智能技术为核心的新技术融合[2]；李广建教授等对智慧情报服务的概念及要实现智慧服务必要研究的知识融合技术等几个重要问题进行了前瞻研究与分析[3]；苏新宁教授提出未来图书馆将会发展成为一种网络运行的数字化、虚拟化、可移动、智慧服务的新形态[4]。

结合上述相关前沿研究观点，智慧知识服务已经成为图书文献情报领域关注的焦点与热点，其涉及的核心要素有空间、数据、知识、技术、平台、用户及生态，本文在智慧知识服务的应用实践方法上进行了探索，进而在上述研究的基础上，进一步界定了智慧知识服务的概念，即智慧知识服务是充分利用AI+大数据的信息技术搭建智能文献情报系统，能够让科技情报工作成为灵活运转的以智能文献情报系统为核心的“数据清洗厂”“信息加工厂”“知识生成厂”与“决策制定厂”，使科技情报工作能够快速洞悉变化、凝练问题、聚焦目标、形成解决方案，极大地弥补人类智能上的不足，增强人们应对复杂问题与任务的能力。

* 本文系中国科学院文献情报能力专项“文献情报”数据湖及“数据质量体系建设”(项目编号：Y9290904)研究成果之一。

作者简介：钱力(ORCID：0000-0002-0931-2882)，信息系统部主任，研究馆员，硕士生导师；刘细文(ORCID：0000-0003-0820-3622)，主任，研究员，博士生导师，通讯作者，E-mail：liuxw@mail.las.ac.cn；张智雄(ORCID：0000-0003-1596-7487)，副主任、中国科学院武汉文献情报中心主任，研究馆员，博士生导师；刘会洲(ORCID：0000-0002-7808-8570)，主任，研究员，博士生导师。

收稿日期：2020-11-05 修回日期：2021-02-17 本文起止页码：78-90 本文责任编辑：杜杏叶

下面将从业界在智慧知识服务方面的具体进展分析、中科院文献情报中心在智慧知识服务的建设思路、智慧知识服务的生态体系框架设计及应用实践等方面进行分析与介绍。

2 当前知识服务面临的问题及业界重要发展方向分析

2.1 知识服务当前面临的主要问题

大数据时代每天产生海量的科技大数据,用户的需求更趋于个性化、定制化以及扁平化的发展趋势。张晓林教授提出颠覆性变革与后图书馆时代将推动知识服务的供给侧结构性改革[5],直接反映出我国科技资源以及科技情报供给与需求不平衡的问题成为当前面临的主要矛盾,而且由于用户与科技情报资源的严重不对称导致该矛盾成为面临的核心问题,迫切需要利用先进数据技术与服务平台解决[6]。

其中,科技知识资源供给与需求不平衡主要体现在:科技知识资源检索发现、海量知识资源却精准主动推送给用户、数据价值没有得到充分挖掘、知识计算引擎没有设计并启动、科技知识流动生态环境没有有效形成、特色数据/专题数据/科研实体知识资源无法快速供给、应急专题/科技界急迫解决知识创新需求所需要的科技知识资源供给速度缓慢等系列问题;而科技情报供给与需求不平衡主要体现到:由于以知识计算为核心引擎的科技大数据中心仍未能有序建立起来的背景下,面对应急、专题、常规等不同特征的科技情报需求时,仍然以人工为主开展数据源遴选、数据收集、数据分析与报告撰写,导致科技情报服务响应速度相对较慢,处理科技情报服务任务数量非常有限;"大数据 + 大平台 + 专家智慧"的工程化情报服务模式由于受到数据及平台的限制,仍然没有有效形成。

2.2 业界重要发展方向分析

面向上述数据、平台及情报服务发展面临的供给侧不平衡问题,业界在理论及工程方面,也有相关探索及应用,充分利用大数据与 AI 技术,在情报服务模式、知识服务范式及知识挖掘方面,利用文本深度学习、结构化分析、知识对象挖掘与结构聚类,发现科学研究中的关于具体方法、过程、参数和结果等的研究设计指纹,支持对解决方案的挖掘和对比分析[7],情报分析走向智能计算的趋势越来越明显[8],文献情报知识服务正面临重大发展机遇。

(1)科技情报机构将 AI 技术作为开展科技情报的核心手段。IARPA 自 2011 年至今连续部署 AI 相关项目,利用"机器智能 + 专家智慧"高度融合的混合智能情报分析模式,通过 AI 技术学习海量科学数据,快速发现科学知识与潜在的科学假设。相应的技术直接应用到开放创新产品中,如 Polyplexus[9]能够提出研究假设、生成创新想法,同时提供创新想法可行性研讨、论证以及市场孵化的创新环境,创建了一个开放网络环境下的问题求解或科技创新的新路径。

(2)专业出版机构借助数据优势,利用 AI 技术进行数据增值,并推出了新型知识服务。如 Taylor & Francis 利用机器学习开发了知识图谱工具 Wizdom.ai[10],数据总量达 150TB,提供了知识计算型的全价值链条的情报智能分析服务模式;Digital Science[11]从研究人员、科研机构、基金项目与出版物这 4 大数据维度,研发智能工具,面向科研全流程,创造一种科研信息服务新模式;Elsevier[12]研发了 Scopus、HiveBench、Mendeley、Pure、SciVal、Funding Solutions、Expert Lookup 以及 Analytical Services 等数字化、知识化工具,有效满足了科研人员的科研需求,基本覆盖了从数据、证据、工具与智慧服务的新型科研生态。

(3)传统的学术评价分析方法基于 AI 技术得到突破创新。如 Semantic Scholar[13]利用 AI 技术对学术文献语义内容进行计算分析,自动识别"现代最有影响力的生物医学研究者"[14],发现人才评价新模式,同时从文本中挑选出最重要的关键词和短语,而不依赖于作者或出版商的键入;它还能够帮助科学家理解论文的内容,这正是谷歌搜索引擎有待提高的地方;与此同时,它还可以找出论文所引用的真正具有影响力的参考文献;使用 AI 来帮助用户筛选大量的科学论文,并在一定程度上理解检索到的科学论文的内容。

(4)AI 技术在专业领域的细粒度知识挖掘方法得到应用。在新材料[15-18]、化学[19-22]、物理[23-25]等基础科研领域,借助 AI 技术创造了面向语义内容智能识别与知识计算的情报分析服务的新模式;智慧知识湖[26]的方案也逐步成熟应用,促进知识密集型业务的形成;Semanic Scholar 提供智能学术搜索引擎、智能影响力评价;Entellect[27]将药物、靶点和疾病数据进行整合,以 AI 方式帮助生命科学公司提供计算服务;BenchSci[28]实现了搜索速度比传统的人工筛选提高了 24 倍,挑选抗体的文献成本减少了 75%;IRIS. AI[29]通过人工智能帮助企业研发部门或高校的研究人员来进行学术论文的筛选;学术出版商 Springer Nature 出版了第一本由机器学习生成的书籍[30];Yewno[31]模仿人类大

脑的运作方式,通过全文分析、计算机语义分析来提炼文献的含义。

(5)以 Bert 预训练模型的新型 NLP 技术为知识计算提供新方法。由于大数据的出现,以及算力的强大支持,深度学习语言模型技术也逐步成熟并得到实际应用,机器翻译几乎接近人类水平[32];特别以 Bert 预训练模型为新型自然处理技术的主流技术在知识抽取及情报计算方面的应用效果比传统方法有明显提升,如全领域的预训练模型 SCIBert[33] 以及专业领域的预训练模型 BIOBert[34],国内哈工大、百度等针对开放的中文资讯型数据训练了模型并发布[35-36],中科院文献情报中心也研发了以 CSCD 中文数据集为核心的中文科技预训练模型[37]。

由上可见,科技文献情报服务领域确实正面临重大发展机遇,但同时也具有重大挑战,在上述相关先进技术方法及应用中仍然存在一些不足需要攻关与进一步突破,例如:IARPA 相关成果在 AI 技术应用结果的语义可理解性及过程的可解释性,仍然需要投入更多的专家智慧干预与参与;专业出版机构在相关智能应用推出的同时,如何保障数据更新的及时性、开放性以及数据质量的精准性,同时如何嵌入到用户科研空间中,为未来开放科研生态做的提前准备,特别是对于相关封闭固守的出版商更要未雨绸缪知识服务新生态架构;同时人工智能技术,特别是以语言预训练模型的深度学习方法发布,对于传统自然语言处理技术是一个重大突破,利用大数据与强大的算力,在专业领域知识深度理解方面,已经初见成效,但如何让深度学习模型嵌入领域的专业知识、预训练模型根据应用场景不同而进行预训练目标的优化及改进,进而实现学习的特征与模式更专业、更精准与更全面等仍需要进一步研究与实破。

面对上述发展机遇及重大挑战,本文主要从两大方面开展研究设计与应用实践:一方面充分利用大数据与 AI 技术带来的实质突破性的优势,建立了文献情报科技大数据中心,即文献情报数据湖,并且研究研发基于科技文献的人工智能引擎,为文献情报的数据增值及智能化应用提供新动力;另一方面主要是面向未来开放科学的重大发展趋势,构建开放数据生态,以“一个数据中心、一个数据中台与一个用户认证体系”的平台化、网络化、协同化模式,实现数据发布、知识共享、知识发现、思想交流、成果管理及情报分析的集成服务。

3 智慧知识服务生态体系框架设计

面向上述分析结果以及中国科学院文献情报中心用户与问题需求,本文设计了智能化、精准化的智慧知识服务生态体系架构,比起传统知识服务系统,其具有的特色及优势如下:

(1)重构数据化与知识化的图书文献情报大数据建设模式,通过打通用户与用户、用户与知识、知识与知识的链接通道,进而构建一个开放数据生态,实现不同用户角色(学术研究、管理服务及决策制定等)、参与并贡献不同阶段的数据服务生态。

(2)强化支撑智慧知识服务的智能文献信息服务平台与工具体系建设,强化面向科技管理决策、学科领域发展和经济社会发展的情报研究服务体系建设,实现面向多来源多场景的科技情报服务场景下,以工程化与工具化的思路,加快知识提炼速度、提升情报响应效率。

(3)设计由“一个数据中心、一个数据中台、一个用户认证体系”支撑的智能化、协同化的智慧知识服务生态体系,建立开放、链接的数据可持续发展机制,通过提供“大平台、小工具”的思路一方面有效集成嵌入到用户的科技情报服务过程中,另一方面,也建立了中国科学院自主知识产权的科技情报平台,保障了科研人员的科研活动信息及科技大数据的自主可控、信息安全。

3.1 设计思路与方法

将大数据和 AI 技术作为智慧知识服务生态体系建设的新引擎与新动力,以数据驱动的思路创建支撑科技创新与发展的“科创知识库”,充分利用 AI 技术搭建智能情报系统,让科技情报工作成为灵活运转的以智能情报系统为核心的“数据清洗厂”“信息加工厂”“知识生成厂”与“决策制定厂”,这一过程使科技情报工作能够快速洞悉变化、凝练问题、聚焦目标、形成解决方案,极大地弥补人类智能上的不足,增强人们应对复杂问题与任务的能力。

3.2 生态体系框架设计

基于上述设计思路与研究方法,本文设计了以“科情大脑”为指挥中心的智慧知识服务生态体系的总体框架,如图 1 所示,即打造了覆盖从科技管理与科技决策、中科院科技创新、科技创新系统其它单元以及社会学术信息环境的全生命周期的数据流、知识流的开放型生态体系。该体系围绕“科情大脑”,构建智慧数据(即科技文献与科技知识大数据中心)、智慧中台(即

知识计算平台与工具体系）与智慧服务（即面向不同应用场景的并基于微服务的智慧知识服务平台）三大智慧知识服务平台，基于智慧中台，灵活面向全生态体系的多需求场景及多用户问题的服务需求提供多样化与个性化的服务功能，以下是详细设计思路：

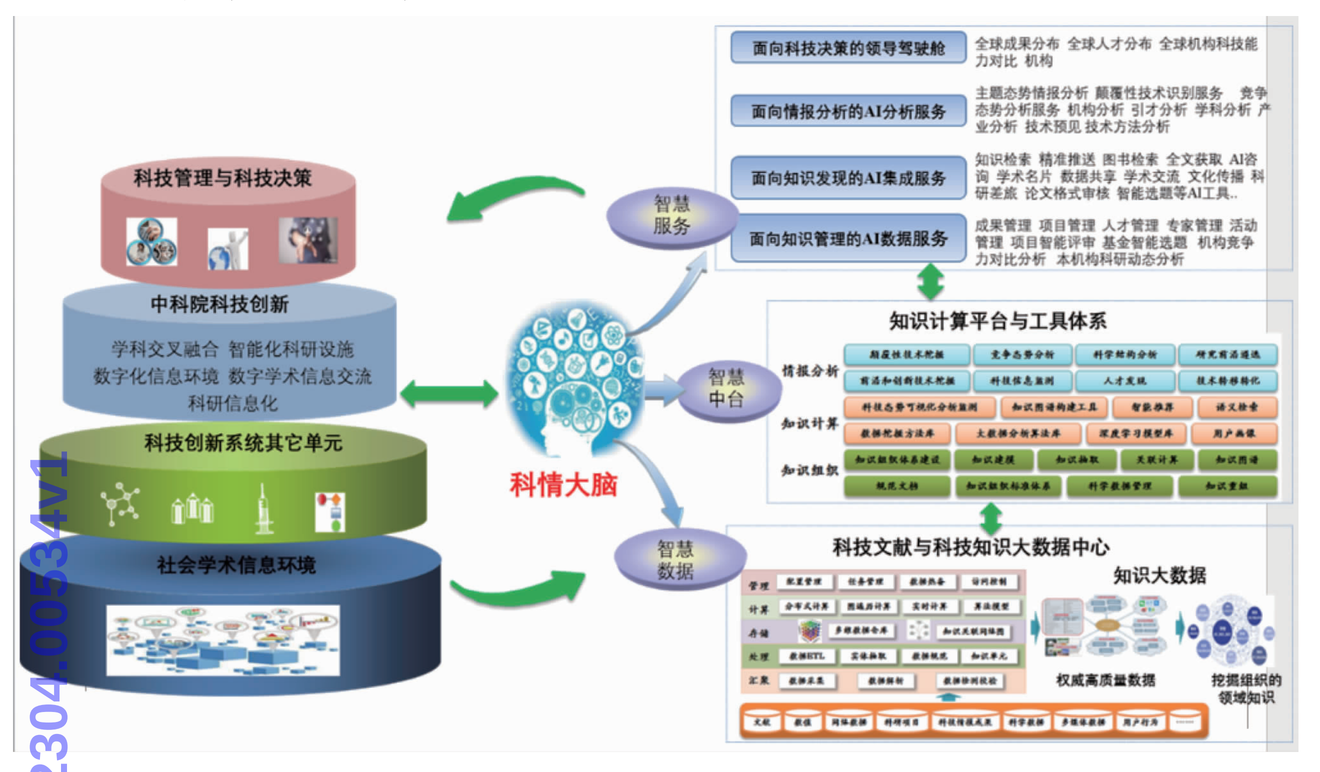


图 1 智慧知识服务生态体系框架

3.2.1 智慧数据研究设计

面向建设支撑科技创新的国家级“科创知识库”的目标，并支持转型升级到以知识计算型为核心的数据服务，本文的智慧数据架构设计了科技文献基础数据库与知识大数据库（科创领域知识库与科创知识图谱），为智慧中台的深度学习模型训练、知识组织计算提供不同层次与深度的智慧数据服务。

（1）构建覆盖全面、权威及时的科技文献基础数据库，即科创基础数据库（见图 2）：从科研主体（专家学者、科研机构、学术期刊、科研团队、出版平台、科技企业、资助机构）、科研活动（科研项目、学术会议、培训交流、科技大赛、数据分享、新闻资讯、社交活动、科技政策）、科研成果（论文、专利、报告、获奖、专著、标准、软件、产品、数据）、科研装置（大科学装置、仪器设备、耗材制剂、研究方法等）及科学数据（研究数据等）五大维度构建了“科创基础知识库”，实现汇聚融合，并从学科分类、产业分类、主题分类、STKOS（科技知识组织体系）范畴分类进行深度标引，对于知识分类计算提供了基础高质量数据。

（2）基于内容挖掘识别细粒度知识智能构建科创领域知识库：基于科创基础数据库，利用 BERT 预训练

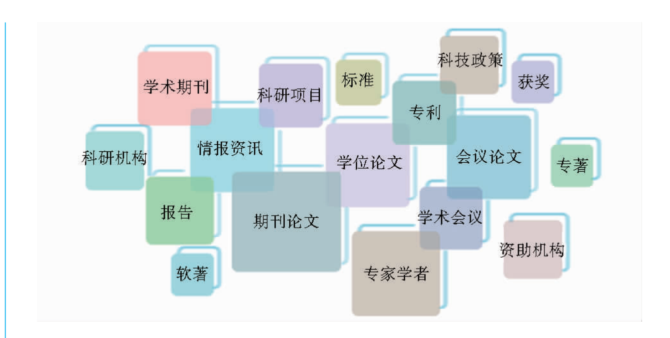


图 2 科创基础数据库已建设数据类型

模型新型 NLP 技术方法，在人工智能与化学键能两个领域，分别示范构建领域知识图谱，其中人工智能领域的知识库包括研究问题、研究方法、研究数据及实现指标 4 类细粒度知识；化学键能领域的知识库包括化合物、溶液、方法、PKA、PKA-VALUE、Bond 与反应 7 类细粒度知识。

（3）基于知识关联计算智能构建科创知识图谱（见图 3）：在科创基础数据库的基础上，制定了数据融合与关联的规则，利用大数据与人工智能技术，对多源异构科技资源进行治理融合、关联计算，联通了各个创新主体与创新资源实体（论文、期刊、学者、机构、项目、主题等），已经建成了知识关系种类有 21 种、知识关系总量

100 亿 + 的科创知识图谱。“科创知识库”实现了“智慧数据”突破传统专家智库型的决策支撑模式,以知识计算的机器智能模式为解决科学问题提供科学办法;科创知识库作为支持数据密集型创新的必要基础设施,可对创新要素进行知识计算、关联推理与深度挖掘,进而促进数据共享、深化数据应用,对我国的科技创新将会产生积极的推动作用;科创知识库的建设,对于我国推动数据共享、深化数据应用、支撑数字经济发展具有重大意义。目前,科创知识库在人才识别、机构评价、项目评估、技术分析、创新发现等方面已经发挥重要的示范作用。

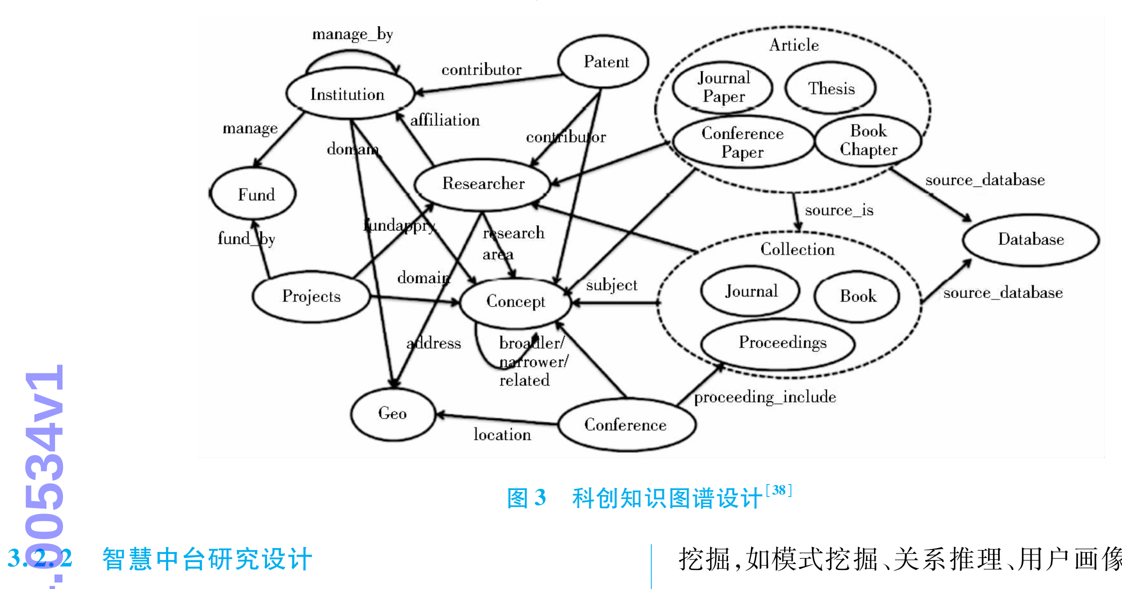


图 3 科创知识图谱设计[38]

3.2.2 智慧中台研究设计

智慧中台的建设目标是建设新一代知识智能技术工具平台,提供丰富的知识智能分析工具,并通过微服务进行整合和封装,统一向上层多场景、多需求的智慧服务应用提供技术支持和数据知识输出,从而提高知识服务产品的建设质量和效率。主要从三种类型来研究设计:

(1)以知识组织工具为核心的中台服务,面向全院知识服务大数据中心体系中的各类资源和数据,建立知识组织应用体系,加强语义增强的技术方法、工具体系,实现对数据的知识内容进行语义化丰富化揭示和组织,提供标引、自动分类、数据关联、知识汇聚、关联查询、知识重组等应用服务,释放数据资源的知识价值和知识组织工具的应用价值:知识抽取工具,如语义标注、实体抽取、关系抽取、事件抽取、主题词提取、上下文关系识别等;知识组织体系的构建、加工和管理工具,如分类管理、叙词表管理、本体管理、RDF 转换、关联数据发布等;知识融合工具,如实体对齐、概念对齐、实体链接、关联挖掘等;知识图谱的构建工具等。

(2)以知识计算工具为核心的中台服务,结合区块链、智能挖掘、社会计算、深度学习等技术,封装知识挖掘和智能分析相关工具和模型,通过松耦合方式构建面向具体业务的应用系统,通过快速配置资源和接入现有技术,加快知识计算应用系统的构建速度,降低人力和资源成本,为业务快速创新提供坚实保障:知识挖掘,如模式挖掘、关系推理、用户画像、社区计算、团队识别等;智能分析,如语义检索、智能推荐、预测可视化、技术挖掘、主题演化等。

(3)以专业领域知识组织与情报计算分析为核心的中台服务:专业化的、垂直型知识服务是支撑科技创新的必要途径,聚焦领域特性,拓展和凝练面向具体专业领域或数据资源的工具,如针对领域实体或概念的抽取和识别,探索面向科研过程的专业知识组织工具,提供领域知识融合服务,发现与开发专业领域相关的技术、方法、工具,更有效地支持科研人员的科研活动,提供更完善的领域战略布局、竞争力分析、发展态势分析、研究机会发现、研究热点分析等知识服务。

3.2.3 智慧服务研究设计

以数据与服务场景驱动的智慧服务设计理念,实现关键知识服务功能组件的无缝融合与嵌入,支持智能导航与引导用户使用需求,一方面提供主动发现用户的需求问题,智能推荐;另一方面提供用户可以按照个性化需求,在不同的应用场景中探索知识,解决需求问题。具体主要面向四类用户角色,提供四大应用场景,但场景之间是统一认证、数据资源一体化管理、同步消息机制,保障在一定程度上,按照用户联想思维,连续提供知识服务。详细功能实现效果参见 4 应用实践部分。

(1)面向知识管理的 AI 数据服务,即面向科研机构对象,实现知识成果的主动精准分发、精准机构画

像、实时机构情报分析，提供机构画像、成果管理、项目管理、人才管理、专家管理、活动管理、项目智能评审、基金智能选题、机构竞争力对比分析、机构科研动态分析等。

(2)面向知识发现的 AI 集成服务，即面向公共用户的知识智能检索发现服务，提供知识检索、精准推送、图书检索、全文获取、AI 咨询、学术名片、数据共享、学术交流社区、文化传播、智能工具(科研差旅、论文格式审核、智能选题、项目评估、科技查新等)等。

(3)面向情报分析的 AI 分析服务，即面向情报分析人员的数据管理与情报智能分析服务，提供主题态势情报分析、颠覆性技术识别服务、竞争态势分析服务、机构分析、引才分析、学科分析、产业分析等。

(4)面向科技决策的领导驾驶舱，即面向决策人员提供全球科研成果的动态扫描与感知分析服务，提供全球成果分布、全球人才分布、全球机构科技能力对比分析等。

4 应用与实践

基于本文设计的智慧知识服务生态体系架构，中国科学院文献情报中心在"十三五的科技知识服务转型升级"上进行了全面实践应用，特别在数据自动采集、自动汇聚融合、智能知识计算、精准服务以及多维度数据画像等方面取得了明显应用成效，为图书文献情报方向在支持知识服务上，奠定了坚实的数据、技术与平台基础。其中，主要在六个方面取得了阶段性应用成果：首先构建了"文献情报数据湖"及"文献情报知识服务引擎工具"，从智慧知识服务的底层提供了丰富了数据知识与算法工具；其次研发了面向知识发现的智慧知识服务平台、面向知识管理的机构数字资产管理与分析平台以及专题情报数据管理与智能分析平台，为普适型的知识发现、知识管理与情报分析提供了工具平台；最后设计了面向科技决策的领导驾驶舱。具体介绍如下：

4.1 实现 AI +"文献情报数据湖"与知识服务设施建设

构建了覆盖全球的科技大数据知识资源，形成了面向科技创新的"科创知识库"，包括全球科技文献、专利数据、科技人才、项目数据、图谱数据、监测数据等，同时建立了国家数字资源长期保存中心[39]，实现国外 65 个资源、国内 3 个资源的全文本地化保存，为科技创新提供了战略资源保障。

其中特色功能包括：①"文献情报数据湖"(见图 4)的数据治理与计算智能化程度达到 90% 以上，实现了机构名称智能规范、智能分类、智能摘要、关键词智能抽取等；②建立了人机融合的规范、精准、结构化治理的"数据湖"云治理服务平台[41]，实现了 19 种数据实体的在线实时治理；③提供快速构建专题数据库功能，支持从文献情报"数据湖"中按照预设的专题知识结构，快速提取并汇聚成权威专题数据库，如新型冠状病毒专题 - 知识服务与科研攻关交流平台[42]；④利用大数据与 AI 技术，构建了包括 4 亿 + 数据集记录、10 亿 + 科技实体、100 亿 + 知识关系的学术知识图谱[43]，形成了能够支撑科技创新的国家级科创知识库；⑤建成了 200 个行业领域的知识库[44]，支撑情报监测与科技决策；数据安全、软件安全及平台安全，在当前面临重要的问题，在我们文献情报"数据湖"及知识设施的建设上，数据分布式存储、分布式计算、分布式索引等基础设施软件都以自主研发为主。

4.2 构建 AI + 文献情报知识服务引擎与工具

基于"文献情报数据湖" + AI 技术，充分利用科技文献大数据集的优势，运营 BERT 预训练模型，研发了科技文献人工智能服务引擎。

其中特色功能包括：①基于文"文献情报数据湖"海量的元数据，利用深度学习技术，训练了系列知识计算服务引擎[45-46]，部分截图参考图 5 和图 6，包括文献分类、关键词识别、概念句子识别、文本标签生成、审稿人推荐、命名实体识别、技术与问题的智能识别、实体名称智能规范等方面，支持了语义挖掘研究和探索发现，该成果 2020 年 12 月 5 日对外正式发布；②利用上述的知识服务引擎，进一步应用基于"预训练 + 微调"模式的 NLP 预训练模型的深度学习技术，在专业领域知识图谱的构建上，形成了方法体系，并在人工智能领域[46]、化学键能领域的细粒度知识识别抽取，形成领域知识图谱，例如能够智能识别人工智能领域的研究问题、研究方法、研究数据与研究指标的细粒度知识，面向文献全文进行了化学键能科学数据自动识别与数据库构建。

4.3 研发了 AI + 面向知识发现的智慧知识服务平台

实现了从多元数据聚合、知识集成检索发现、智能综述、主题分析、知识精准主动推送、学术名片及服务自动导航(下载/文献传递/参考咨询/主题分析/图谱发现等场景)等系列功能，以及支持 PC、微信及 APP 多终端的一体化服务及运营体系，即形成文献情报服务的主服务门户[47]，已经于 2020 年 11 月 11 日更新上线，根据 2020 年 12 月份数据显示，日均访问量是旧服务系统的 3 倍多，见图 7。

图 4 文献情报“数据湖”数据治理平台[40]



图 5 基于科技文献知识资源的 AI 引擎



图 6　面向专业领域的细粒度知识挖掘工具

图 7　中国科学院文献情报中心智慧知识服务门户(http://www.las.ac.cn)

其特色功能包括：①基于学术知识图谱的多类型知识的统一检索与关联发现，从论文、专利到报告文献类数据统一检索，到学者、机构、主题类数据的统一检索，再到一篇文献关联到主题、学者、文献、机构的关联发现（正在增加相关数据的关联发现），突破传统只提供该篇文献元数据展示功能；②提供了对海量检索结果的智能综述功能，快速了解检索结果知识；③计算出检索对象的“语义相关主题”，能够按照某一主题进行智能分析，快速了解该主题的发展态势及年度核心论文等情报信息；④提供了基于用户 ID 认证下载服务集成功能，支持用户随时随地按照权益下载论文全文；⑤无缝集成了“慧科研 – 智能随身科研助理”所有功能[48]（见图 8），特别是基于用户画像与 AI 技术，提供了“自动创建学者学术名片、个人学术成果校验与管理、高价值知识的自动推送、个性化知识订阅、开放数据共享、以及用于前沿问题快速求解及思想火花碰撞交流的创新社区”等特色功能，也提供支持基金项目指南发布的智能立项选题、科研差旅、科技查新、申报项

目的智能评估、论文格式的智能审核等智能功能，努力打造科技界的今日头条；无缝集成了中科院文献情报中心的其他工具平台及专家团队、数据资源，实现了一个平台的一站式资源发现与获取；平台研发提供了用户统一认证中心，不仅仅支持用户可以基于 ID 账号按照所在单位订购权益下载全文，而且实现了在主服务系统中跨系统使用一次登录认证即可。



图 8　慧科研－智能随身科研助理

4.4　研发了 AI＋面向知识管理的机构数字资产管理与分析平台

基于“文献情报数据湖”，利用大数据与 AI 技术，在机构名称智能规范、科研成果智能精准分发、人名智能规范等关键技术环节突破的基础上，以数据驱动的理念，研发了面向知识管理的机构数字资产管理与分析平台（https://inst. scholarin. cn/），对机构进行全方位、多视角的分析评估，辅助机构管理者管理机构知识资产（见图 9）。同时，该平台还为本机构科研成果认定、人员身份信息审核等方面提供了客观数据支撑，一定程度上在学术道德规范方面提供了一个有效的数据分析平台。

其特色功能包括：①实现了科研成果按照机构维度进行智能精准分发功能，自动构建了机构科研知识资源数据库，突破传统由人工逐条数据上传的数据管理方法；②提供实时机构全景画像功能，从机构的科研成果产出趋势、合作机构网络、高产出论文学者、高产出专利发明人、高产出项目负责人以及论文收录数据统计、年度研究热点主题等维度进行实时的全景画像；③提供科研队伍管理，从职称、年龄、性别、部门团队、研究方向、专家团队等方面实时分析，并提供在线数据化管理功能；④提供科研项目全流程管理功能，从项目申报、审批、开题、中期、结题全流程的数据化管理服务；⑤提供科研成果分析功能，实时感知与分析机构的科研成果类型分布、产出趋势、SCI 及其他收录实时数据统计、检索发现等功能；提供支持本地化科研成果知识资产的一体化管理功能，按照数据规范格式，批量上传本地化的知识资产，实现一体化管理，从而盘活本地化知识资产；提供机构订购数据库使用数据管理以及使用用户授权管理。

4.5　以数据驱动的思路建设了专题情报数据管理与智能分析平台

基于“文献情报数据湖”，利用大数据与 AI 技术，以“数据 ＋ 平台 ＋ 专家”的混合智能模式，研发了“AI＋面向情报分析的智能分析平台”（http://ai. scholarin. cn/）（见图 10），尝试解决数据分散、无法积累重用、无法共享、快速分析以及提供数据服务的问题。

其特色功能包括：①以平台化的方式，支持快速融入专家智慧，实现专题情报方向的知识结构快速构建，为后续的情报分析获取更精准、全面的分析数据集打下基础；②面向一个专题情报分析场景，提供本地数据集及规范数据（国家、机构、人名及关键词等）导入到平台中，实现线上与线下数据的一体化管理，达到数据



图 9　慧科研 – 机构知识管理与分析服务平台

可积累、可重用及集中管理的目标；③基于文献情报数据湖的规范库数据，提供在线数据自动清洗的功能；④提供基于传统科学计量学的统计分析服务功能，包括论文数据的发展态势分析、机构分析、地域分析、期刊分析、关键词分析、热点分析；专利数据的发展态势分析、技术分析、专利权人分析、发明人分析、地域分析、关键词分析与研究热点分析；基金项目数据的发展态势分析、资助单位分析、负责人分析、承担机构分析、地域分析、关键词分析、研究热点分析；⑤探索尝试基于语义计算的面向内容分析的服务：研究问题分析、关键技术分析、问题与技术关联分析以及研究热点技术识别；提供快速生成“数据型情报分析报告”，辅助情报人员进行情报分析，加快情报生产速度与对科研情报需求的响应速度。



图 10　慧科研 – 专题情报数据管理与智能分析平台

4.6　创建了 AI + 面向科技决策的领导驾驶舱

设计与创建了决策分析沉浸式感知环境（见图 11），融合了文献情报数据湖、智能知识引擎及可视化等核心技术，能够进行图像识别、智能语音交互、情报可视化分析、科技情报动态感知等方面提供更直观、更便捷、更精准、更智慧的环境，为面向科技决策以及科技情报分析研判提供了新模式与新方法。



图 11 面向科技决策的沉浸式科技情报智能分析场景

5 结论与展望

当前,大数据与人工智能技术为行业经济发展带来了重大发展机遇,同时也为科技知识服务服务模式提升带来了变革升级的机会,同时也遇到诸多挑战。本文在上述大的背景下,综述了当前大数据与 AI 在知识服务方面的应用实践,结合文献情报领域的发展时机,提出了构建 AI + 智慧知识服务的生态体系架构,在数据汇聚、知识计算、工具研发、知识服务平台研发等方面进行了深入研究设计与实践,并公开发布了 6 大智慧知识服务系统,面向中国科学院、全国省级科学院以及部分研究机构进行了实际应用,得到用户的普遍好评,为未来文献情报知识服务生态形成奠定了较好基础。未来,该研究将继续围绕 AI + 智慧知识服务生态体系架构,在数据的精度、知识的深度、服务的专业度、情报的及时度以及智慧知识服务体系的智能化程度方面,继续完善与提升智慧知识服务能力及水平。

致谢:本文研究成果得到文献情报能力专项突破一的大力支持,特别感谢相关参工作成员人员,感谢全中心同事的支持,感谢信息系统部相关产品设计、数据计算、系统开发及网络支撑人员。

参考文献:

[1] 张冬荣, 钱力. "科技大数据与智能知识服务平台建设"专题序[J]. 数据分析与知识发现, 2019, 3(1): 3-3.
[2] 柯平,邹金汇. 后知识服务时代的图书馆转型[J]中国图书馆学报. 2019,45(1):4-17.
[3] 罗立群,李广建. 智慧情报服务与知识融合[J]情报资料工作. 2019,40(2):87-94.
[4] 苏新宁. 新时代图书馆使命与未来图书馆学教育之思考[J]. 中国图书馆学报,2020(1):53-62.
[5] 张晓林. 颠覆性变革与后图书馆时代-推动知识服务的供给侧结构性改革[J]. 中国图书馆学报, 2018, 44(1): 4-16.
[6] 钱力,谢靖,常志军,等. 基于科技大数据的智能知识服务体系研究设计[J]. 数据分析与知识发现, 2019, 3(1): 4-14.
[7] 钱力, 张晓林, 王茜. 基于科技文献的研究设计指纹描述框架研究[J]. 大学图书馆学报, 2015, 33(1): 14-20.
[8] 李广建, 江信昱. 论计算型情报分析[J]. 中国图书馆学报, 2018, 44(2): 4-16.
[9] About us-polyplexus. com [EB/OL]. [2020-10-06]. https://start. polyplexus. com/about-us/.
[10] wizdom. ai - intelligence for everyone[EB/OL]. [2020-10-06]. https://www. wizdom. ai/#about.
[11] Digital science [EB/OL]. [2020-10-06]. https://www. digital-science. com/.
[12] Research intelligence [EB/OL]. [2020-10-06]. https://www. elsevier. com/research-intelligence.
[13] A free, AI-powered research tool for scientific literature[EB/OL]. [2020-10-06]. https://www. semanticscholar. org/.
[14] Who's the most influential biomedical scientist? computer program guided by artificial intelligence says it knows | Science | AAAS [EB/OL]. [2020-10-06]. http://www. sciencemag. org/news/2017/10/who-s-most-influential-biomedical-scientist-computer-program-guided-artificial.
[15] TSHITOYAN V, DAGDELEN J, WESTON L, et al. Unsupervised word embeddings capture latent knowledge from materials science literature[J]. Nature,2019, 571(7763), 95-98.
[16] KIM E, HUANG K, SAUNDERS A, et al. Materials synthesis insights from scientific literature via text extraction and machine learning[J]. Chemistry of materials,2017,29(21): 9436-9444.
[17] ZHOU Q, TANG P, LIU S, et al. Learning atoms for materials discovery[J]. Proceedings of the National Academy of Sciences, 2018,115(28): E6411-E6417.
[18] KIYOHARA S, MIYATA T, TSUDA K, et al. Data-driven approach for the prediction and interpretation of core-electron loss spectroscopy[J]. Scientific reports, 2018 8(1): 1-12.

[19] STOKES J M, YANG K, SWANSON K, et al. A deep learning approach to antibiotic discovery[J]. Cell, 2020, 180(4): 688-702.

[20] SEGLER M H, PREUSS M, WALLER M P. Planning chemical syntheses with deep neural networks and symbolic AI[J]. Nature, 2018, 555(7698): 604-610.

[21] GRANDA J M, DONINA L, DRAGONE V, et al. Controlling an organic synthesis robot with machine learning to search for new reactivity[J]. Nature, 2018, 559(7714): 377-381.

[22] POPOVA M, ISAYEV O, TROPSHA A. Deep reinforcement learning for de novo drug design[J]. Science advances, 2018, 4(7): 7885.

[23] RADOVIC A, WILLIAMS M, ROUSSEAU D, et al. Machine learning at the energy and intensity frontiers of particle physics[J]. Nature, 2018, 560(7716): 41-48.

[24] MATHURIYA A, BARD D, MENDYGRAL P, et al. CosmoFlow: using deep learning to learn the universe at scale[C]//SC18: international conference for high performance computing. Dallas Texas: IEEE, 2018: 819-829.

[25] ZHANG Y G, GAJJAR V, FOSTER G, et al. Fast radio burst 121102 pulse detection and periodicity: a machine learning approach[J]. The astrophysical journal, 2018, 866(2): 149.

[26] BEHESHTIL A, BENATALLAN B, SHENG Q Z, et al. Intelligent knowledge lakes: the age of artificial intelligence and big data[C]//WISE 2020. Singpore: Springer, 2020.

[27] Connecting life science data to enable impactful analytics - Entellect | Elsevier [EB/OL]. [2020-10-06]. https://www.elsevier.com/solutions/entellect.

[28] AI-assisted reagent selection and experiment design- BenchSci[EB/OL]. [2020-10-06]. https://www.benchsci.com/.

[29] Your science assistant[EB/OL]. [2020-10-06]. https://iris.ai/.

[30] 学术出版商 Springer Nature 出版了第一本由 AI 作家创作的书籍[EB/OL]. [2020-10-06]. http://m.elecfans.com/article/906694.html.

[31] Transforming information into knowledge[EB/OL]. [2020-10-06]. https://www.yewno.com/.

[32] Google's neural machine translation system: bridging the gap between human and machine translation[EB/OL]. [2020-10-06]. https://arxiv.org/pdf/1609.08144.pdf.

[33] SciBERT: a pretrained language model for scientific text[EB/OL]. [2020-10-06]. https://arxiv.org/abs/1903.10676.

[34] BioBERT: a pre-trained biomedical language representation model for biomedical text mining. [EB/OL]. [2020-10-06]. https://arxiv.org/abs/1901.08746.

[35] Pre-training with Whole Word Masking for Chinese BERT[EB/OL]. [2020-10-06]. https://arxiv.org/abs/1906.08101.

[36] 中文任务全面超越 BERT:百度正式发布 NLP 预训练模型 ERNIE[EB/OL]. [2020-10-06]. https://zhuanlan.zhihu.com/p/59436589.

[37] 基于科技文献预训练模型[EB/OL]. [2020-10-06]. http://sciengine.las.ac.cn/.

[38] 王颖,钱力,谢靖,等. 科技大数据知识图谱构建模型与方法研究[J]. 数据分析与知识发现,2019(1):15-26.

[39] 国家数字科技文献资源长期保存体系[EB/OL]. [2020-10-06]. http://ndpp.ac.cn/.

[40] 科技大数据多元知识融合治理云平台. [EB/OL]. [2021-02-16]. http://data.scholarin.cn.

[41] 科技大数据多元知识融合治理云平台[EB/OL]. [2020-10-06]. http://data.scholarin.cn/.

[42] 新型冠状病毒专题-知识服务与科研攻关交流平台[EB/OL]. [2020-10-06]. https://ncov.scholarin.cn/.

[43] 大数据探索发现系统[EB/OL]. [2020-10-06]. http://kgview.las.ac.cn/discover.

[44] 领域科技情报监测服务云平台[EB/OL]. [2020-10-06]. http://stmcloud.las.ac.cn/.

[45] 基于科技文献知识资源的 AI 引擎[EB/OL]. [2020-10-06]. http://sciengine.las.ac.cn/.

[46] 深度语义挖掘工具[EB/OL]. [2020-10-06]. http://finger.las.ac.cn/Page_sharingTool_tm_NER.html.

[47] 中国科学院文献情报中心知识服务门户[EB/OL]. [2020-10-06]. http://www.las.ac.cn.

[48] 慧科研-智能随身科研助理[EB/OL]. [2020-10-06]. http://scholarin.cn.

作者贡献说明:

钱力:论文的撰写与修改,智慧知识服务平台的研发实现;

刘细文:AI+智慧知识服务生态体系架构设计,论文内容的修改与指导;

张智雄:文献情报数据湖及知识服务引擎的建设与指导,论文内容的修改与指导;

刘会洲:提出 AI+智慧知识服务平台构建内容,并对论文内容的修改与指导。

Design and Application of Ecological System of Intelligent Knowledge Service Based on AI
——An Example of Building of Intelligent Service Platform of National Science Library, CAS

Qian Li[1,2]　Liu Xiwen[1,2]　Zhang Zhixiong[1,2]　Liu Huizhou[1,2]
[1] National Science Library, Chinese Academy Sciences, Beijing 100190
[2] Department of Library Information and Archives Management,
University of Chinese Academy of Sciences, Beijing 100190

Abstract: [**Purpose/significance**] Artificial Intelligence is triggering a chain reaction-like scientific breakthrough, leading a new round of scientific and technological revolution and industrial transformation, how to use AI to provide Intelligent Knowledge Services and Intelligent Information system is the current focus of attention and hot spot. [**Method/process**] This paper analyzed AI technology and big data from inside and outside the library and information domain to bring new platforms, new services and new opportunities and challenges to the knowledge service paradigm, and provided the idea of building the Intelligent Knowledge Service ecosystem driven by "AI technology and big data", and built the "science brain" method from the three levels of intelligent data, wisdom center and intelligence service, and provided an open intelligent knowledge service ecological environment covering science and technology management, scientific and technological innovation and social academic information environment. [**Result/conclusion**] As an exploration building of National Science Library Chinese Academy of Sciences, about Data Lake of Library and Information, Intelligent Knowledge Service engine, Intelligent Knowledge Discovery, Intelligent Knowledge Management, Intelligent Knowledge Analysis and Intelligent Sense Environment, and good results have been achieved. In the future, AI technology for big data government, fine-grained knowledge recognition, precision service and so on, still need to be further improved in data, technology and service models.

Keywords: artificial intelligence-ai　sci-tech big data　intelligence knowledge service　deep learning

《图书情报工作》杂志社发布出版伦理声明

为加强和增进学术论文写作、评审和编辑过程中的学术规范、科研诚信与学术道德建设,树立良好学风,弘扬科学精神,坚决抵制学术不端,建立和维护公平、公正、公开的学术交流生态环境,《图书情报工作》杂志社(包括《图书情报工作》《知识管理论坛》两个期刊编辑部)结合两刊实际,特制订出版伦理声明并于 2020 年 2 月正式发布。

该出版伦理声明承诺两刊将严格遵守并执行国家有关学术道德和编辑出版相关政策与法规,规范作者、同行评议专家、期刊编辑等在编辑出版全流程中的行为,并接受学术界和全社会的监督。共包括三大部分,总计十五条,分别为:一、作者的出版伦理(①学术论文是科学研究的重要组成部分;②学术不端是学术论文的毒瘤;③作者是学术论文的主要贡献者;④ 作者署名体现作者的知识产权与学术贡献;⑤学术论文要高度重视知识产权与信息安全;⑥参考文献的规范性引用是学术规范的重要表征;⑦要高度重视研究数据与管理的规范性;⑧建立纠错与学术自我净化机制)。二、同行评议专家的出版伦理(⑨同行评议是论文质量的重要控制机制;⑩评审专家应遵守论文评审的相关要求;⑪评审专家要严格遵循相关的伦理指南和行为准则)。三、编辑的出版伦理(⑫编辑应成为学术论文质量的守护者;⑬编辑应在学术道德建设中发挥监控作用;⑭编辑要成为遏制学术不端的最后屏障;⑮对学术不端实行"零容忍")。

全文请见:http://www.lis.ac.cn/CN/column/column291.shtml

(本刊讯)